

ARI Contractor Report 2008-02

**Measuring Learning and Performance in Collective
Training Exercises**

**David H. McGilvray, Bruce C. Leibrecht,
and Karen J. Lockaby**
Northrop Grumman Technical Services

This report is published to meet legal and contractual requirements and may not
meet ARI's scientific or professional standards for publication.

March 2008

**United States Army Research Institute
for the Behavioral and Social Sciences**

Approved for public release; distribution is unlimited.

20080424048

**U.S. Army Research Institute
for the Behavioral and Social Sciences**

**A Directorate of the Department of the Army
Deputy Chief of Staff, G1**

Authorized and approved for distribution:


BARBARA A. BLACK, Ph.D.
Research Program Manager


MICHELLE SAMS, Ph.D.
Director

Research accomplished under contract
for the Department of the Army

Northrop Grumman Technical Services

Technical review by

Bruce Knerr

NOTICES

DISTRIBUTION: Please address correspondence concerning distribution of reports to:
U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-
ZXM, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This Contractor Report may be destroyed when it is no longer
needed. Please do not return it to the U.S. Army Research Institute for the Behavioral
and Social Sciences.

NOTE: The findings in this Contractor Report are not to be construed as an official
Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yyyy) March 2008		2. REPORT TYPE Final		3. DATES COVERED September 2006 - November 2007	
4. TITLE AND SUBTITLE Measuring Learning and Performance in Collective Training Exercises				5a. CONTRACT OR GRANT NUMBER W74V8H-04-D-0045 (DO 0014 and DO 0023)	
				5b. PROGRAM ELEMENT NUMBER 622785	
6. AUTHOR(S) David H. McGilvray, Bruce C. Leibrecht, and Karen J. Lockaby (Northrop Grumman Technical Services)				5c. PROJECT NUMBER A790	
				5d. TASK NUMBER 294	
				5e. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Northrop Grumman Technical Services 2011 Sunset Hills Road Reston, VA 20190				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: DAPE-ARI-IF 2511 Jefferson Davis Highway Arlington, VA 22202-3926				10. MONITOR ACRONYM ARI	
				11. MONITOR REPORT NUMBER ARI Contractor Report 2008-02	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES Contracting Officer's Representative: Bruce W. Knerr This report is published to meet legal and contractual requirements and may not meet ARI's scientific or professional standards for publication.					
14. ABSTRACT (<i>Maximum 200 words</i>): The goal of the research described in this report was to develop a proof-of-principle scoring system that can be used to evaluate training effectiveness across diverse scenarios. The focus was on supporting evaluators as they evaluate and track unit performance across scenarios. The report describes the products of the research as well as the insights and lessons learned. A scoring system with a computer interface suitable for a hand-held computer was developed and tried out with Infantry subject matter experts acting as evaluators observing virtual scenarios. The try-out provided empirical data on the utility of the scoring system and on desired improvements. Based on feedback from the try-out, the scoring system was revised. The report contains findings and lessons learned that can guide future efforts to automate evaluator and Observer/Controller (O/C) support tools.					
15. SUBJECT TERMS					
Scoring Systems Infantry Squad Evaluation Collective Training		Evaluation Tools Small Unit Performance Training Evaluation		Team Performance Hand-Held Computer Tools Unit Performance Measurement	
SECURITY CLASSIFICATION OF					
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	19. LIMITATION OF ABSTRACT Unlimited	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON Ellen Kinzer, Technical Publication Specialist, 703.602.8047

ARI Contractor Report 2008-02

**Measuring Learning and Performance in
Collective Training Exercises**

David H. McGilvray, Bruce C. Leibrecht, and Karen J. Lockaby
Northrop Grumman Technical Services

ARI-Orlando Research Unit
Stephen L. Goldberg, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

March 2008

Army Project Number
622785A790

Personnel Performance
and Training Technology

Approved for public release; distribution is unlimited.

ACKNOWLEDGMENTS

The research described in this report could not have been conducted without the combined efforts of numerous people, in addition to the report's authors. The Northrop Grumman team members whose efforts were critical to this report include:

- Richard L. Wampler, who was instrumental in developing the revised scoring system and served as try-out coordinator.
- Dr. Jack Hiller for his invaluable contributions to the revised scoring system.
- Mike Dover and George Mabry, who developed the initial version of the competencies.
- Sean M. Cooley and Tony N. Fullen, developers of the user-computer interface software.

We thank the Northrop Grumman subject matter experts who participated in the try-out to evaluate the scoring system and computer interface developed during this research – Jim Centric, Mike Dlubac, and David James. Their expertise yielded both a thorough assessment of the tool's utility and valuable insights for the improvement, expansion, and application of this evaluation approach.

Finally, thanks go to Dr. Stephen L. Goldberg whose conceptual input played a key role in shaping the scoring system.

MEASURING LEARNING AND PERFORMANCE IN COLLECTIVE TRAINING EXERCISES

EXECUTIVE SUMMARY

Research Requirement:

The U.S. Army Research Institute (ARI) has long conducted research in the application of training technology to collective training. In this research, ARI has found that measurement of learning during training of small units using multiple scenarios has been hampered by the circumstance that units seldom repeat exactly the same scenarios. This requires the Soldiers to apply lessons learned during one scenario while executing subsequent scenarios. Inevitably, the question arises: Did the training intervention produce learning? There is a lack of a conceptual framework for measuring and interpreting unit performance independent of the specific scenario's conditions. To investigate this conceptual issue, the current research effort aimed to develop the foundation for a tool that can be used to evaluate unit performance, and thereby training effectiveness, in a scenario independent manner. To this end, the research focused on developing, demonstrating, and refining a general-purpose scoring scheme for evaluating small unit training performance.

Procedure:

This research was conducted in multiple stages culminating in the try-out of a prototype infantry small unit scoring system. In the initial task, a literature review identified design principles relevant to performance assessment support tools. In parallel, infantry subject matter experts (SMEs) developed and iteratively refined small unit competencies, along with tasks and measures that are scenario independent. Concurrently, a prototype computer interface suitable for a hand-held computer was designed and developed. For the try-out, the prototype scoring and interface systems were integrated and tested using infantry SMEs who were not previously involved in this research. Based on feedback from the try-out, the research team made major revisions to the small unit competencies.

Findings:

The try-out provided empirical data on the utility of the scoring system and on desired improvements. A qualitative analysis of the feedback from the infantry SMEs indicated that the core competencies/tasks, based on the Army's primary training standard, were not fully suited to application across scenarios. Therefore, the team revised the scoring system by creating ten scenario-independent tasks. The structure of the tasks stemmed from the Army's established plan-prepare-execute-consolidate/reorganize phases of mission accomplishment, and a theory-based model of command and control information processing. The report describes the revised system and offers recommendations and lessons learned for follow-up research.

Utilization and Dissemination of Findings:

This research establishes an innovative springboard for designing and developing methods and tools for evaluating collective performance in operational and simulation-based training. The products and findings offer practical help to follow-on investigators in the area of unit training evaluation. The insights and lessons learned will help researchers, working in concert with SMEs, to fully develop a performance measurement system that is scenario independent and applies to multiple unit types and echelons. By creating this foundation, the findings move toward a general-purpose performance measurement system capable of charting a unit's proficiency improvement across diverse training events.

MEASURING LEARNING AND PERFORMANCE IN COLLECTIVE TRAINING EXERCISES

CONTENTS

	Page
INTRODUCTION	1
Background	1
Technical Objectives	2
METHOD	3
Overview	3
Literature Review	3
Development of Tasks, Supporting Behaviors and Scoring Scheme	3
Development of Automated Interface	4
Try-Out	7
Revision of Scoring System	11
RESULTS AND DISCUSSION	12
Performance Design Principles	12
Core Competencies	13
Infantry Small Unit Scoring System	14
Try-Out Results	17
Revised Scoring System	20
Lessons Learned	25
CONCLUSIONS AND RECOMMENDATIONS	28
Conclusions	28
Recommendations	30
REFERENCES	31
Appendix A. Acronyms and Abbreviations	A-1
Appendix B. Data Collection Instruments	B-1
Appendix C. Literature Review References	C-1
Appendix D. Details of Core Competency Tasks as Used During Try-out and Subsequently Rejected	D-1

CONTENTS (Continued)

Appendix E. Details of Revised Infantry Squad Evaluation Criteria and Scoring E-1

List of Tables

Table 1.	<i>Design Goals for the User Interface</i>	5
Table 2.	<i>Structural Elements of the Evaluator Tool</i>	5
Table 3.	<i>Functions Represented in the Evaluator Tool</i>	6
Table 4.	<i>Targeted Characteristics of the User Interface</i>	7
Table 5.	<i>Try-Out Plan Summary</i>	8
Table 6.	<i>Try-Out Schedule</i>	9
Table 7.	<i>Literature Review Lessons Learned</i>	13
Table 8.	<i>Summary of the Nine Core Competency Tasks with Steps</i>	14
Table 9.	<i>Summary of Scoring System and Interface Components</i>	15
Table 10.	<i>Tasks Used by SMEs</i>	19
Table 11.	<i>Revised Competencies with Supporting Actions</i>	23
Table 12.	<i>Critical Lessons from Literature Review</i>	29

List of Figures

Figure 1.	System Map Screen.....	16
Figure 2.	Example of Task Assessment Screen	17
Figure 3.	Hiller's C2 Information Processing Model.....	22

MEASURING LEARNING AND PERFORMANCE IN COLLECTIVE TRAINING EXERCISES

INTRODUCTION

To support the United States Army's training efforts, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) investigates training needs. One critical need is the capability to effectively evaluate collective training performance independent of a specific scenario. In the research project entitled *Measuring Learning and Performance in Collective Training Exercises (MLPCTE)*, the ARI Orlando Research Unit set out to develop a scenario independent scoring system that would allow a single subject matter expert (SME) or trainer to evaluate the performance of a small unit as it conducts a series of dismounted infantry scenarios. The intended result of this process was the capability to track changes in a unit's performance over the course of multiple training scenarios. This report describes the products and lessons learned from the MLPCTE project. The findings and products point the way to research that will provide a useful training evaluation tool for assessing the effectiveness of Army training interventions.

Background

There are numerous Army evaluation programs for unit collective performance. The evaluation programs, such as those found in Army Training and Evaluation Programs (ARTEPs) and Training Support Packages (TSPs), are normally associated with scripted scenarios which prompt the performance of the desired skills. Following each exercise, an after action review (AAR) is conducted as a feedback session to provide the Soldiers a learning opportunity focusing on what they did, how well they performed, and how to improve their performance the next time. However, the unit does not normally repeat the execution of the exact scenario that was previously performed, partially to prevent advance knowledge of scenario events, and partially because changes in unit behavior can cause each occurrence of a specific scenario to play out differently.

In conducting research on collective training and training technology, ARI has found that measurement of learning during training of small units across multiple events has been hampered because units seldom repeat scenarios. Rather, units routinely use multiple scenarios, calling for performance under altered conditions or requiring the performance of different collective tasks. This means that Soldiers must apply lessons learned during one scenario while subsequently executing different scenarios. Inevitably, the question arises: Did a training intervention produce learning? The lack of a conceptual framework for measuring and interpreting unit performance independent of the specific scenario's conditions results in the inability to reliably answer this question.

To investigate this conceptual issue, the goal of this research was to establish the foundation for a tool that can be used to evaluate training effectiveness in a scenario independent manner. To this end, the focus was on: (a) development of a scoring scheme which is scenario independent, usable by a single evaluator for evaluating a small unit as it conducts a training exercise, and capable of tracking unit performance across multiple scenarios; (b) automation of

the scoring scheme for use on a hand-held computer; (c) the try-out of these applications in a controlled research environment; and (d) revision of the scoring scheme based on the results of the try-out. The intent of this effort was to determine the functionality of the materials developed and to gain insight into the progression, expansion, and application of this evaluation approach to further develop its potential as a training tool during future work.

The importance of this effort was reinforced by the Chief of Staff of the Army (CSA) in his article in the 2007 edition of the Army Green Book (Casey, 2007) in which he stressed the critical requirements for unit and leader training, especially in the current operating environment. Key to his guidance was the necessity of training to deal with a broad range of missions across the spectrum of conflict and the necessity of being able to perform those missions globally as required by an ever changing world situation. The essence of this training requirement is scenario independent training.

Technical Objectives

As the foundation for developing a performance measurement system that is scenario independent and that has future application to other unit types and echelons, the current report concentrates on the development of the prototype scoring system that constitutes the first step toward that goal. This will provide the foundation for researchers to further develop this type of performance measurement in future projects. The following technical objectives, as refined during the execution of the project, guided the research described in this report:

- Review existing collective performance measurement schemes for applicability to measuring learning and performance in small unit Army operational scenarios.
- Analyze Army collective tasks for competencies that are scenario independent and establish a set of core competencies for dismounted infantry small unit operations.
- Develop a performance measurement scoring system with user computer interface that enables a single SME or trainer to evaluate the performance of a small unit across multiple training events.
- Try out the scoring system using a series of prerecorded scenarios with qualified SMEs to determine the efficacy and utility of the prototype system and to gain insight for developing its potential during future research.
- Based on the results of the try-out, recommend revisions to the scoring system, including the core competencies.

METHOD

Overview

The goal of this project was to develop and demonstrate a prototype performance measurement system, assess its functionality, and gain insight into the refinement and application of the approach. This required that the research be accomplished in distinct stages:

- Development of a scenario independent scoring system based on dismounted infantry squad core competencies.
- Development of a computer interface suitable for use on a hand-held computer.
- Try-out of these conceptual products in a controlled research environment.
- Revision of the performance measurement system based on results of the try-out.

The research approach combined military subject matter expertise, behavioral science knowledge, and computer programming to execute the four stages. The research team relied primarily on military training assessment principles, a thorough review of Army documents and websites relating to core competencies applicable to infantry dismounted small unit operations, and the expertise of infantry SMEs working in collaboration with a computer technology expert to develop the scoring system and the computer interface. Simple formative evaluation methods were used to try-out the utility and efficacy of the scoring system developed during this research. Finally, the initial scoring system was revised based on the try-out results.

Literature Review

The initial stage of this research effort entailed a review of existing performance measurement schemes for applicability to this project. A number of research projects that have developed measures of collective performance were reviewed, including those identified in a search of ARI and Defense Technical Information Center libraries. Researchers with experience in the field of collective training performance measurement also identified relevant articles. The review was conducted to determine the applicability of existing performance measurement schemes to the current research and to inform the current work. For this reason, findings from this review were considered lessons learned to be built upon in development of measurement methods under this project. Additionally, information not directly applicable to the current project may be of value to inform future research efforts that further advance, expand, or apply the results of the current work. The review was conducted using a template that, in addition to bibliographic data, collected information on the purpose of the performance measurement scheme; measurement context; and scheme characteristics (tasks measured, structural dimensions, types of metrics).

Development of Tasks, Supporting Behaviors and Scoring Scheme

The development of a set of core competencies for small unit operations was an iterative process which utilized SMEs as key members of the research team. All SMEs had extensive experience as training developers at the Infantry Center and School at Fort Benning, GA, had authored ARTEP mission training plans (MTPs), and had participated in analyzing infantry small unit performance of collective tasks.

The initial step in the analysis to develop core competencies was a review of Army training publications and other sources to establish a list of applicable collective missions and tasks. Materials reviewed included:

- ARTEPs
- MTPs
- Unit Mission Essential Task Lists (METLs) and Tactical Standing Operating Procedures (TACSOPS)
- ARI reports
- Army lessons learned
- Relevant websites

This review and analysis resulted in a list of approximately 60 collective tasks for the dismounted infantry small unit. After analyzing, inventorying, and categorizing high-payoff tasks and supporting behaviors (subtasks) that a unit would perform under a variety of scenarios, an initial list of core competencies was developed. During this analysis, the research team defined core competencies as the base skills in which a small infantry unit must be proficient regardless of the unit's assigned role or mission. Each iteration of the core competency list was reviewed by other research team members not involved in the list development and then revised until the final list received concurrence. Steps and measures were then determined for each core competency task.

Development of Automated Interface

The development of the interface was initiated concurrently with the development of the core competencies used in the scoring system. The software designer collaborated with the infantry SMEs for the integration of the scoring system and interface to facilitate the functionality of the evaluator tool for the try-out phase of the research. The interface design was a sequential process beginning with the determination of design principles and then moving through the stages of developing goals, structural elements, functions of the tool, and determination of the targeted characteristics.

The first step in the development of the interface was to determine the design principles that would underlay its development. Predicated on the characteristics specified in the technical objectives and with amplification through a series of discussions between the design team's infantry SMEs, computer SME, and researchers, the team envisioned a highly portable job assistant with database-supported capabilities. The following principles guided the development of the tool:

- Support efficient operations of evaluators directly observing live and virtual exercises.
- Provide readily mobile, continuous (approximately 2 hrs), stand-alone operations.
- Incorporate essential performance measurement, assessment and review functions only.
- Rely on commercial technology that requires no modification beyond programming.
- Ensure consistency with contemporary training practices and environments.
- Present user-friendly look and feel (Windows), consistent across all functions.
- Emphasize user-controlled navigation with minimal risk of getting lost.
- Ensure continuous user awareness of pathway environment and options.

- Reduce evaluator workload by use of guided measurement and streamlined data entry features.
- Enable a user to perform full operations with minimal (less than 10 min) tutoring.
- Minimize need for external job aids, assessment guides, etc.
- Assume user is tactically and technically qualified to serve as an infantry Observer/Controller (O/C).
- Provide open-ended architecture for future expansion to other echelons and units.

These design principles were then translated into design goals to provide dimensional guidance for the design process, as specified in Table 1.

Table 1.
Design Goals for the User Interface

Dimension	Goal
Target Audience	Evaluators pre-qualified to support training of current Infantry units
Echelon Applicability	Tailored to performance parameters of company, platoon, squad echelons
Operating Model	<i>Stand-alone, fully portable device that interfaces with database (notional)</i>
Domain Versatility	Usable in home station, Combat Training Center, and deployed arenas
Scope of Functions	Performance evaluation activities before/during/between exercises
Customization	Limited to user options available on host platform
Pathway Guiding	Menu-prompted navigation with limited system-controlled guiding
User Alerting	Audible signal when pre-programmed conditions are met
System Activation	Immediate, single-click start-up of software from desktop
Security	<i>Password protection of access to functions and stored data (notional)</i>
Role of Automation	Limited mainly to selective guiding of performance measurement

The components or structural elements were specified along with a short description of the outcomes desired for each element. As stated in the *Background*, the intent of this research included laying the foundation for future development of the scoring system's potential to provide a training evaluation system or tool with application to other unit types and echelons. For this reason, provisions were made for a notional database and interface of the tool with that database. Structural elements of the evaluator tool are specified in Table 2.

Table 2.
Structural Elements of the Evaluator Tool

Element	Description	Desired Outcome
Interface Platform	Lap or hand-held device w/limited input/output options (including touch screen)	Enable evaluator to use anywhere, optimize ease of operation
Function Set	Suite of multi-phase evaluation functions for evaluator use (see Table 3)	Minimize evaluator workload in all supported activities
Database (Notional)	Remote (central) repository for tasks/steps, measures, data, references	Provide one-stop, up-to-date source of guidelines and data
Database Interface (Notional)	Wireless download, upload, and administrative functions	Facilitate transparent linkage between evaluator and database

The use of the scoring system and interface requires that the evaluator understand the functions of the interface and execute those functions during sequential phases. The functions required or desired during each phase are listed in Table 3. Notional functions were specified to facilitate future development.

Table 3.
Functions Represented in the Evaluator Tool

Phase	Functions
System Start-up	<ul style="list-style-type: none"> • Login with user authentication • Select user-controlled options and preferences • Initialize at main menu (automatic)
Orientation	<ul style="list-style-type: none"> • About this tool (purpose, ownership, version, etc.) • System overview (list of functions with brief explanation, instructions) • System map (graphic layout of functions)
Exercise Preparation	<ul style="list-style-type: none"> • Review tasks, steps and measures (via evaluator-paced process) • Select steps and measures for specific exercise • <i>Correlate steps and measures with scenario events (notional)</i> • <i>Review previous performance data for participating unit(s) (notional)</i> • <i>Specify/select alert parameters and conditions (notional)</i>
Exercise Execution and Assessment	<ul style="list-style-type: none"> • <i>Maintain awareness of exercise progress (notional)</i> • <i>Receive alerts of approaching events or conditions (notional)</i> • Enter information identifying the exercise as a unique event • Record measures and comments (with flexible timing and sequence) • Review/verify data from the exercise, and revise as necessary • Aggregate and organize quantitative data (menu driven) • Review exercise data and record conclusions (evaluator driven) • <i>Perform AAR and feedback functions (notional)</i>
Post-Exercise Activities	<ul style="list-style-type: none"> • <i>Archive exercise data package (notional upload to central database)</i> • <i>Review data from previous exercise (mock-up)</i> • <i>Re-assess performance data from previous exercise (notional)</i> • <i>Derive performance trends from stored data (menu driven) (notional)</i> • <i>Archive performance trends (open-ended for later updating) (notional)</i> • <i>Set or revise access privileges for stored data and trends (notional)</i>
Supporting Activities	<ul style="list-style-type: none"> • <i>Update task/subtask components of database (notional)</i> • <i>Consult training reference materials (e.g., TACSOPs, MTPs) (notional)</i> • <i>Search for information stored in database (notional)</i> • <i>Share archived information with unit trainers (notional)</i> • <i>Perform housekeeping functions with database contents (notional)</i>

To further guide the interface design, desired characteristics of the interface system were developed and described, as delineated in Table 4. These characteristics were intended to facilitate the design of a “user-friendly” interface.

Table 4.
Targeted Characteristics of the User Interface

Characteristic	Description
Windows framework	Use of windows techniques to organize functions and information
Screen utilization	Full-screen windows to avoid keyhole and crowding effects
Display simplification	Minimal text and graphic elements present on each screen
Window contents visibility	Minimal scrolling for visibility of window contents to facilitate viewing
Information streamlining	Avoidance of prosaic verbiage in favor of intuitive, condensed forms
Text entry context	Comment entry option accessible in display of performance measures
Organizational simplicity	Flat schema (single level) for primary functions to simplify navigation
Menu format	Drop-down menus (single click) with highlighted default as appropriate
Awareness aiding	Always visible main menu; menu item highlighting; you-are-here device
Button previewing	Pop-up window summarizing function(s) accessed by every button
Stylistic consistency	Consistent use of colors, shapes, font, emphasis techniques, etc.
Multi-session management	Automatic book marking and "you are here" cueing at end of session
Background cognizance	Status of background functions (e.g., download, upload) visible to user

Try-Out

The purpose of the try-out was to evaluate the utility and efficacy of the small unit scoring system and interface, or evaluator tool. The strategy called for applying the tool in scenario-driven exercises to obtain feedback on the tool's performance, modifications needed, and potential for development during future research. The try-out was conducted in a research environment using pre-recorded virtual exercises, with each exercise independently scored by three infantry SMEs who were not previously involved in the research.

The simulation environment utilized ARI's Dismounted Infantry Virtual After Action Review System (DIVAARS) software (Knerr et al., 2003) to play the recorded exercises. One desktop computer served as the DIVAARS workstation. Scenario materials (described below) came from ARI's library of virtual exercises (Knerr et al.). Playback of a scenario's data stream portrayed a short (about 20 min) combat vignette, including voice communications. The tactical actions could be viewed from optional perspectives. Desktop computers, one for each of three infantry SMEs, served as evaluator workstations to run the scoring system and display the user interface. No data capture capabilities were utilized beyond those built into the scoring system software. Table 5 summarizes the try-out plan.

Table 5.
Try-Out Plan Summary

Aspect	Plan
Objective	Demonstrate the evaluator tool and gather feedback on its suitability, acceptability, potential for development in future research, etc.
Staffing	<ul style="list-style-type: none"> • Three infantry SMEs who were not previously involved in the project • One research team member serving as coordinator and evaluator • One DIVAARS operator
Questions of Interest	<ul style="list-style-type: none"> • How suitable is the tool for assessing collective performance of infantry squads? • How well does the tool enable an evaluator to collect data and assess performance? • How operator-friendly and effective are the scoring system and user interface? • How useful is the tool for tracking task proficiency over multiple scenarios? • How can the scoring system and user interface be improved?
Materials and Equipment	<ul style="list-style-type: none"> • Working copies of the prototype system loaded on desktop computers • Squad-level virtual scenarios in multimedia form • DIVAARS workstation for displaying the virtual scenarios with voice communications
Test Conditions	<ul style="list-style-type: none"> • Dedicated office for a controlled research environment • DIVAARS workstation created a virtual simulation environment • Desktop computers represented evaluator workstations running MLPCTE software
Test Phases	<ul style="list-style-type: none"> • Set-up and verification (one day of testing and resolving problems/issues) • Familiarization and train-up (half-day of orientation and workstation practice) • Scoring of scenarios (two half-days of exercise scoring and data collection) • Post-exercise data collection (half-day of reviews, hotwash, brainstorming) • Wrap-up (administrative actions were completed within 3 working days)
Set-up	<ul style="list-style-type: none"> • The DIVAARS operator and an SME tested functionality of each workstation • Set-up participants fully implemented three scenarios (one cycle per workstation) • Hardware and software problems were resolved • The try-out coordinator verified readiness to execute exercises and collect data
Execution Procedures	<ul style="list-style-type: none"> • The try-out coordinator controlled the flow of events • The SMEs practiced during trial exercises to become proficient on the evaluator tool • The SMEs observed each virtual exercise simultaneously but independently • Each SME recorded performance data on his own evaluator workstation • The SMEs could revisit part of a scenario using a different vantage point • All SMEs completed exercise scoring before moving on to the next scenario • A short break (approximately 5 min) separated one exercise from the next • The coordinator discouraged interaction among the SMEs during exercises

Table 5.

Try-Out Plan Summary (continued)

Data Collection	<ul style="list-style-type: none"> • SMEs recorded unit performance data using the evaluator workstations • The evaluator wrote his observations on a structured data capture form • Each SME completed a paper worksheet following every exercise • The evaluator and SMEs participated in a hotwash following each session • The evaluator recorded and compiled notes from all discussion sessions • Data recorded on the evaluator workstations was exported to compact disc (CD) • The evaluator and SMEs documented lessons learned
Data Handling	<ul style="list-style-type: none"> • Compiled scoring and observation data using Microsoft Office® tools • Analyzed observations, scores, worksheets, and notes for trends and insights • Derived lessons learned regarding measurement methodology and the evaluator tool
Outcomes	<ul style="list-style-type: none"> • Feedback on suitability, acceptability, effectiveness, and value of the tool • Recommendations for improving the user interface and scoring methodology

As seen in Table 6, the try-out was conducted in phases during a three-day period. Phases included a one day pre-execution phase for set up and verification of hardware and software functionality, a four-hour phase for SME train-up on the system, two separate four-hour sessions of scoring and data collection while SMEs observed scenarios, and a post-exercise data collection phase.

Table 6.

Try-Out Schedule

Day	Phase / Activity
Day 1	Set-up and Verification
Day 2	Familiarization and Train-up (half day)
Day 2 and 3	Scoring Sessions I and II (half day each)
Day 3	Post-Exercise Data Collection (half day)

During every exercise, each SME observed the tactical events/actions on the DIVAARS screen and independently recorded performance data using their assigned workstation. The try-out coordinator recorded observations, SME comments, and administrative data using the observation guide. If any SME requested, the DIVAARS operator replayed a segment of a scenario from a different vantage point. At the end of every exercise each SME completed a worksheet. The SMEs were allowed to raise questions and issues between scenarios, but interaction among the SMEs was discouraged while they were scoring an exercise.

Staffing for the set-up event consisted of one DIVAARS operator and one infantry SME, who served as the try-out coordinator. During the verification event, three infantry SMEs were added to the staffing with the evaluator tool developer (computer programmer) available to correct any problems identified. This staffing was maintained through the Familiarization and Train-up and the Post-Exercise Data Collection phases of the try-out.

The three SMEs were retired infantry Soldiers with extensive experience in training squads in both live and virtual training environments. Two were retired noncommissioned

officers (NCOs), a Sergeant First Class (E-7) and a Sergeant Major (E-9). Among the SMEs, leadership experience included positions as squad leader and platoon sergeant; one also served as an infantry company First Sergeant. The third SME was a retired officer, Major (O-4) with prior enlisted experience including positions as squad and section leader as an NCO, and platoon leader and company commander as an officer. All three were veterans with combat experience in Vietnam, Operation Urgent Fury, or Operation Iraqi Freedom. All were instructor certified, with their Army instructor experience including Ranger School cadre and training center Drill Sergeant assignments.

The SMEs had significant experience in a virtual environment including: scenario development; squad mentoring; role-playing various positions to interface with simulation subjects via radio transmissions; development of performance scoring measures and tools; portraying exercise opposing force (OPFOR); and conducting AAR sessions. All three SMEs also had extensive experience in evaluating squad performance during simulation exercises and experience working with several simulation programs.

Scenarios Used for the Try-Out

From the library of nine pre-recorded scenarios produced in previous ARI research (Knerr et al., 2003), five were implemented in a total of seven exercises. Each scenario involved a dismounted infantry squad conducting urban operations. The scenario missions included: Deliberate Attack Version 1 (two iterations), Deliberate Attack Version 2 (one iteration), Hostage Rescue Version 1 (two iterations), Crowd Control, and Downed Helicopter. Each simulation-based scenario portrayed a dismounted infantry squad preparing and executing an assigned mission at a simulated Military Operations on Urban Terrain (MOUT) site. The squad was part of an infantry battalion attached to the United Nations (UN) Protection Force. The latter was conducting operations in a town of strategic importance and opposing rebel forces from a radical nationalist group that was linked to terrorist bombings and attacks on nearby towns.

Data Collection Instruments

The focus of data collection efforts was to gather information on the functionality and applicability of the tool developed during this research and to gain information that might contribute to the progression of this training evaluation approach during future research. Data collection forms are briefly described below and shown in Appendix B.

SMEs Worksheet: This instrument contained ten questions regarding the SMEs' reactions to operating the scoring system and interface during each scenario. Questions also addressed the applicability of the tasks and subtasks to the unit's actions during the scenario. A worksheet was completed at the end of every scenario.

Observation Guide: This form contained 31 questions covering all train-up and scoring activities. The first seven questions gathered information on the SMEs' orientation and train-up. A dozen of the questions were repeated in two sets, one for each of the half-day sessions during

which multiple scenarios were scored. The try-out coordinator used this form to record administrative information (scenario name, start/stop time, replay aspects) and observations.

Hotwash Guide: This guide contained 34 questions addressing all aspects of the try-out including the adequacy of SME training, doctrinal correctness of tasks and supporting behaviors, and the functionality and utility of the scoring system and interface. The try-out coordinator used the questions to facilitate a hotwash at the conclusion of each half-day session.

Revision of Scoring System

The scoring system was revised based on the SME feedback from the try-out. The revision team consisted of an infantry SME, a retired infantry field grade officer with over 24 years of active duty and extensive research experience, and researchers in the field of Army training evaluation and assessment. The feedback was integrated with training theory and modified and refined through an iterative discourse to focus on core competencies that were scenario independent and potentially applicable to other unit types and echelons.

RESULTS AND DISCUSSION

Performance Design Principles

The initial aspect of this research was the review of prior research with the intent of both informing the current research and providing information that may be applicable to future research projects that investigate developing the potential of this evaluation approach. The review included 26 research reports, articles, book chapters, and conference papers which covered a wide range of research on training within the Army, Air Force, Navy, and a multi-service project. Although many articles had some information to contribute, few constituted “training measurement schemes.” Relevant schemes included:

- Mission Essential Competencies (MECs) (Colegrove and Bennett, 2006; Alliger et al., 2003) based on United States Air Force research on measuring the proficiency of air combat aircrews or other weapon systems operators with the intent of improving training. Although originally designed for use with air crews, Alliger et al. report the application of MECs to a unit, describing the unit in the context of a weapon system.
- A human systems integration method for validating team performance assessment (Johnston, Vincenzi, Radtke, Salter, and Freeman, 2005). This U.S. Navy project used a hand-held tablet computer to assist three separate two-person observer teams in assessing team performance on 14 key, pre-specified training objectives during a scenario, with the observers receiving an alert prior to the occurrence of events.
- The Target Acceptable Responses to Generated Events or Tasks (TARGETS) (Fowlkes, Lane, Salas, Franz, and Oser, 1994; Throne, Holden, and Lickteig, 2000). This Army effort used non-SME observers to record the presence or absence of acceptable individual and team task responses while observing filmed scenarios. Observers worked with a list of events and responses, and received cues when events were about to occur.
- The Observer Assessment Scheme (Kyne, Militello, Thordsen, and Klein, 2002; and Kyne, Thordsen, & Kaempf, 2002) is another Army training measurement scheme focused on team decision-making and performance assessment. The SME observers subjectively rated team performance using 16 behavioral dimensions and a five-point scale (paper form). Also included was an observer support package that served as a quick reference guide – one page per behavioral dimension which included a definition of the dimension, descriptions of indicators of effective performance, and space for observer notes.

During the review of the literature, many lessons were gleaned from research in the field of measuring learning and performance in small unit Army training scenarios. Key lessons learned appear in Table 7, grouped into categories.

Table 7.

Literature Review Lessons Learned

Category	Lesson Learned
Development of Competencies	<ul style="list-style-type: none"> - Detailed analysis is critical, including involvement of SMEs - Analytical process should focus on key (vs. all) tasks or events - Research and real-world factors should be balanced - Development schedule should accommodate multi-phase, iterative process - Mission essential competencies are a high-value approach
Team Framework	<ul style="list-style-type: none"> - Collective assessment hinges on definition of team concept - A team should be viewed as a unitary, intelligent entity - Teamwork evaluations should relate to mission essential competencies - Team performance is more than the sum of parts - Team performance may depend on a single member for some tasks
Performance Measurement	<ul style="list-style-type: none"> - Collective measures should reflect overall unit performance - Performance standards should be bands or ranges, not point values - Measurement scaling can enhance differentiation of performance - Complete assessment integrates network and non-network data - Automated measures can increase scope, objectivity, and precision - At present, automated measures are not wholly sufficient - Feasibility (via automated or evaluator collection) is a possible key criterion - Design of measures should consider output format requirements
Observer Considerations	<ul style="list-style-type: none"> - Capabilities of intended observers should be defined - Highly proficient, motivated observers should be a key priority - Managing observer workload is a central consideration - If multiple observers are used, work allocation becomes important
Data Collection	<ul style="list-style-type: none"> - Automated measurement tools can reorient observer's focus - Automated collection and computation can overload system - Simulation-based recording of exercises can extend analysis - Integrating automated and evaluator measures can be a challenge
Observer Job Aids	<ul style="list-style-type: none"> - Observers must maintain good situational awareness (SA) - Automated support (e.g., event alerting, battle tracking) can be valuable - Automated job aids do not redeem unqualified observers - Ready access to references can enhance observer effectiveness
The Feedback Connection	<ul style="list-style-type: none"> - Presentation of automated results may be a design consideration - Automated tools can compare performance to standards - Pictorial and graphical presentation of measures is difficult to achieve

Core Competencies

Using the ARTEP as the standard for dismounted infantry small unit combat specific tasks yielded a lengthy list of combat tasks. These MTP tasks were then analyzed and reduced to the key and essential tasks, resulting in a list with a usable number of tasks. Nine core competency tasks were selected as a basis for the scoring system. Table 8 summarizes the core competency tasks.

Table 8.

Summary of the Nine Core Competency Tasks with Steps

Task	Steps	
1. Breach an Obstacle	18 Steps	9 Leader, 9 Unit
2. Conduct a Defense	24 Steps	14 Leader, 10 Unit
3. Conduct a Movement to Contact	14 Steps	7 Leader, 7 Unit
4. Conduct a Security Patrol	13 Steps	6 Leader, 7 Unit
5. Conduct Tactical Movement	24 Steps	12 Leader, 12 Unit
6. Conduct an Attack	16 Steps	7 Leader, 9 Unit
7. Maintain Operations Security	6 Steps	4 Leader, 2 Unit
8. Action on Contact	14 Steps	7 Leader, 7 Unit
9. Conduct Troop Leading Procedures (TLP)	9 Steps	9 Leader, 0 Unit

For each of the nine tasks, steps and performance measures that captured the key elements of each task were identified using the ARTEP standards. These elements were included in the scoring system to provide the specific information believed to be needed by the evaluators. All but one task included steps for both the leader and the unit, with tasks having a minimum of six steps and a maximum of 24 steps for a total of 138 steps across the nine tasks. Further, performance measures were specified for each step, with a total of approximately 360 performance measures. Details of these tasks are in Appendix D which contains: (a) a table summarizing the core competency tasks and steps within each task; and (b) details of the core competencies with tasks, conditions, standards, task steps, and performance measures.

Infantry Small Unit Scoring System

The scoring system and interface were developed for use by an SME to evaluate a dismounted infantry unit (targeted at the squad level) using scenario independent competencies. The stand-alone system provides the evaluator the ability to assess the performance of unit collective tasks. For this investigation, the Windows-based system was implemented on a desktop computer. However, the system could be mounted on a fully portable hand-held device.

The program was developed using Adobe Flash Player 9 and the coding was created using Actionscripting 2.0. The program is a self-contained Flash file embedded in a Hyper Text Markup Language (html) page for proper viewing. All input values are compiled into a Flash object which acts as a "cookie" on each computer. This allows the information to be retained in place even if the evaluator closes and re-opens the program. The program operates on any Windows-based computer with Flash Player and Internet Explorer installed. The program was developed using Flash Player 9 and Internet Explorer 7, the latest versions of those programs when the software for the scoring system was developed. The program is self-contained in one flash file and all navigation while using the scoring system must take place within the program. Although the program will run directly from a CD, optimal operation results from saving the files to the hard-drive of the evaluator's computer.

After logging in, the evaluator can select scenario options with standard scoring means on the nine collective tasks. The scoring system's major components (main menu) are:

- *Orientation* gives general information and orients the user to the scoring system.
- *Exercise Preparation* enables the user to review the tasks, steps and measures and to review previous performance data for the unit to be observed.
- *Exercise Execution/Assessment* structures recording of scores and comments in a format that reflects the tasks and steps. Review of data entry is also available.
- *Post Exercise Activities* (currently notional) represent desired capabilities that the user will need to finalize the evaluation and to analyze the data.
- *Supporting Activities* (currently notional) will enable updating of competencies or general evaluator functions, including access to “Reference Materials” such as field manuals.

As indicated in Table 9, each of the main menu selections subsumes a set of optional functions. Access to a function is gained by selecting a main menu item. As developed in this project, many functions are notional which are included to indicate the scope of the desired functions and to convey the system’s potential.

Table 9.
Summary of Scoring System and Interface Components

Main Menu Selections	Subordinate Options Available	Development Stage
Orientation	About this Evaluator Tool	Functional
	System Overview	Functional
	System Map	Functional
Exercise Preparation	Review Tasks, Steps and Measures	Functional
	<i>Select Exercise Measures</i>	<i>Notional</i>
	<i>Verify Measurement Plan</i>	<i>Notional</i>
	Review Prior Exercise	Functional
Exercise Execution/Assessment	<i>Set Alerts</i>	<i>Notional</i>
	<i>Event Cues On/Off</i>	<i>Notional</i>
	<i>Alerts On/Off</i>	<i>Notional</i>
	<i>Register Exercise</i>	<i>Notional</i>
	Record Measures and Comments	Functional
	Verify Data	Functional
	<i>Aggregate Data</i>	<i>Notional</i>
	<i>Draw Conclusions</i>	<i>Notional</i>
	<i>AAR/Feedback</i>	<i>Notional</i>
Post Exercise Activities	<i>Archive New Data</i>	<i>Notional</i>
	Review Prior Exercise	Functional
	Re-assess Prior Exercise	Functional
	<i>Compute Trends</i>	<i>Notional</i>
	<i>Archive Trends</i>	<i>Notional</i>
	<i>Set Data Access</i>	<i>Notional</i>
Supporting Activities	<i>Update Tasks</i>	<i>Notional</i>
	<i>Reference Materials</i>	<i>Notional</i>
	<i>Search Database</i>	<i>Notional</i>
	<i>Share Data or Trends</i>	<i>Notional</i>
	<i>Housekeeping</i>	<i>Notional</i>

There are both system controlled and user controlled navigation functions. The user navigates the system using the tabs at the top of each screen. When the section is active, tabs for each internal function are visible. Also, a system map (Figure 1) is included within the Orientation section to facilitate user orientation to the screen and function navigation.

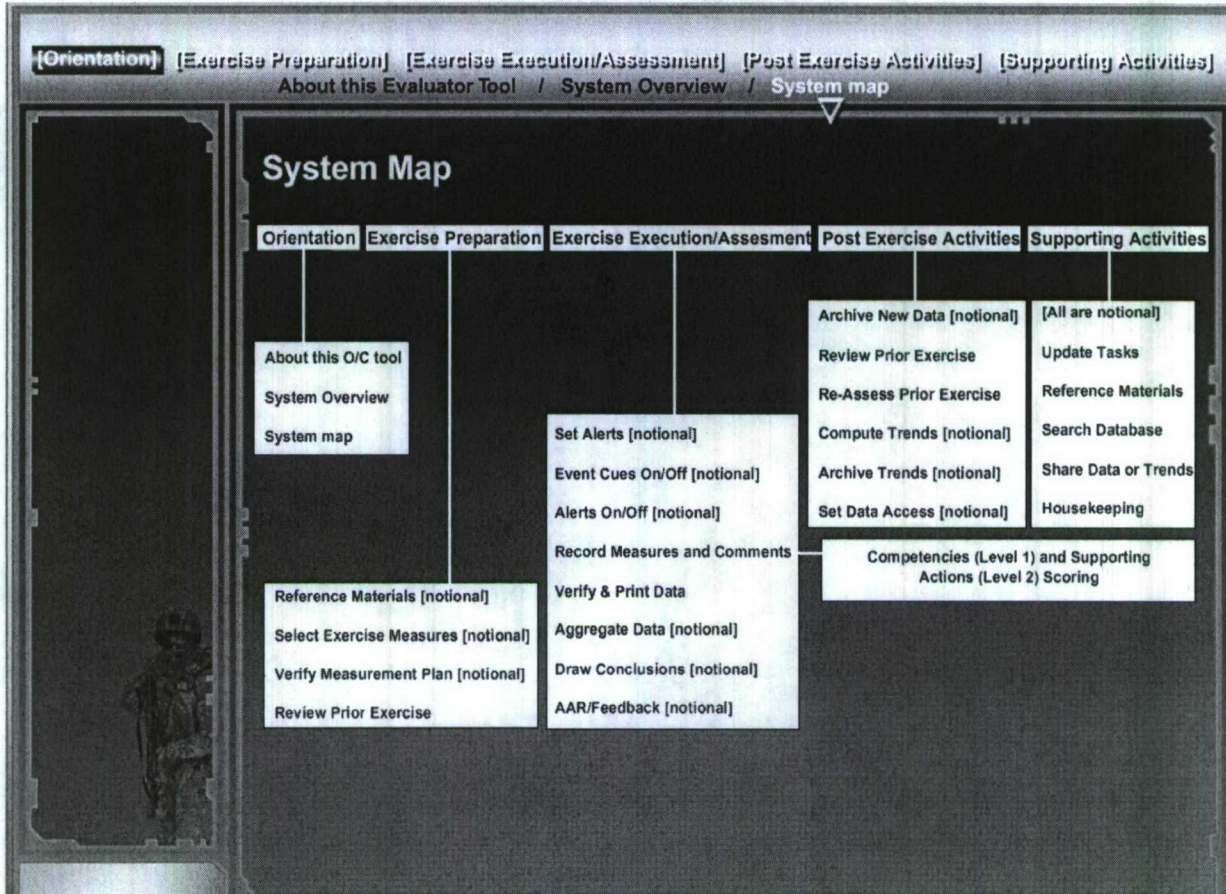


Figure 1. System map screen (Revised Scoring System Version).

This project focused on the capability to use scenario independent competencies to record the assessment of a dismounted infantry squad as it executes a mission. This capability resides in the Record Measures and Comments function within the Exercise Execution/Assessment section.

Detail provided in the scoring system distracted SMEs from observation. The detail provided by steps and performance measures for each task distracted the SMEs from observation during the conduct of the scenarios. The SMEs found it very difficult to devote full attention to the scenario events while they read through the various lists of steps and performance measures (over 300 lines of text) under the scoring system's nine tasks, which were tabbed on their computer screen. They stated that even if they were to become fully familiar with the large volume of steps and measures, they might, at best, only be able to make entries that were most key to their evaluation. There were too many evaluation points for the SMEs to evaluate each one.

The scoring system focused SMEs on task detail rather than scenario independent observations. The detail provided by steps and performance measures focused the SMEs on mission specific details of the squad's performance, rather than scenario independent aspects. While SMEs commented on the excess of material provided and difficulty of finding the proper location to record observations, they were also drawn to the specific details of tasks, steps, and performance measures. They noted other steps and performance measures they believed had been omitted, were included but should be omitted, or suggested the addition of other steps and measures which could be included to address specific squad leader and member actions. The focus was on the detail of a specific sub-task (e.g., sub-tasks required to clear a room in an urban environment), rather than on scenario independent tasks.

The scoring system was not usable for recording data while observing a scenario. The intent was to have the evaluators complete the ratings during the observation of the scenario. Tasks based on the MTP with the numerous steps and measures for each proved to be cumbersome and impractical for real time application during an exercise scenario. During the training phase of the try-out, the SMEs initially attempted to score as they observed the scenario. By the end of the training phase, independently all three SMEs began to make notes on paper during the observation of the scenario, just as they would without an electronic tool. They also stated that as they observed the scenario, they made determinations of the squads' performance without specific reference to the scoring system. They did not actually start making entries in the automated tool until the scenario ended. The SMEs' procedure for making entries was to read through the listings of steps and performance measures to locate the right place to evaluate what they had observed and they had difficulty finding the appropriate place. For the seven scenarios observed, times to complete the scoring after conclusion of the scenario observation ranged from a minimum of 21 minutes to a maximum of 34 minutes. The average time was 29 minutes.

Only a limited number of the tasks were actually used for scoring. Of the nine tasks available for scoring, only four were used at all and only three were frequently used by SMEs (see Table 10). Feedback indicates that in the case of Task 1, Conduct Troop Leading Procedures, the vignettes used for the try-out began after all or most of these actions would have occurred. Therefore, they were not observed by the SMEs. However, the feedback also indicates that the large number of steps and performance measures limited use of the tasks to those that the SMEs determined could be used repeatedly. The SMEs consistently made mental notes of squad performance as they observed, took handwritten notes and then searched the task tabs and lines for a location to enter their evaluation. The feedback also indicated that SMEs

more frequently used familiar tasks, and did not use tasks that they had not previously used in the context of this try-out.

Table 10.
Tasks Used by SMEs

Scenario	Core Competency Tasks								
	Conduct TLP	OPSEC	Tactical Move	Security Patrol	Breach Obstacle	Action on Contact	Move to Contact	Attack	Defend
1			A, C		B	A, C		A, B, C	
2			C		B, C	A, C		A, B	
3			A, B, C			A, B, C			
4			B, C			C		A, B	
5					C	A, C		A, B, C	
6			A, B			B, C		A, C	
7			A, B			A, C		A, B, C	

Note: SMEs designated as A, B, and C

Key to using the scoring system is the degree of user friendliness. Although most of the comments dealing with user acceptability dealt with software issues (discussed below), there are general observations that impact both revision of the current scoring system and future development for operational use. The evaluators stated that the scoring system must be exceptionally user friendly and demonstrate the capability to enable evaluators to perform their duties more efficiently and effectively. Critical aspects to consider are:

- The tool must be extremely user friendly so it can be used like a notepad for recording information and be used while walking.
- The number of rating elements that the evaluator must use must be limited.
- The ability to tailor the scoring system to the scenario being assessed would be beneficial. There were two aspects cited:
 - Arranging the tasks in the order they would appear in the scenario.
 - Selecting or showing only the items that are to be evaluated during the scenario so SMEs do not have to scroll through unused task tabs.
- The ability to use the scores and comments for an AAR would contribute to the system's usefulness.

Doctrinal references are a desirable feature. Although the SMEs became mired in and encumbered by the detail which was displayed on the screens, they recognized the need for doctrinal reference materials. They commented on the need for reference material to use while preparing for evaluator duties. They stated that paper references would be adequate. However, there are advantages to a mature scoring system having electronic references, whether the access is embedded in the system or accessible by hyperlink. The latter method would avoid updating issues as doctrine is changed.

User-Computer Interface

Although all SMEs stated that the interface was easy to understand and use, they made recommendations for improvement. These are categorized and summarized below.

- Visibility of electronic features. All features need to be easily readable by the evaluator, especially if the scoring system is to be used in a field environment. Specific improvements included better color contrast for the cursor and scroll bars.
- Ease of entering scores and recording information. The ability to make entries, including written comments, quickly and easily is essential if evaluators are going to adopt the system for their use. Suggestions included:
 - For a space where a point entry evaluation is to be entered, make the entire box area active, and not just a tiny circle.
 - For use in a field environment, improve text entry options for speed, accuracy, and flexibility (e.g., text entries similar to cell-phone keypad with limited keys, where multiple presses of the same key represent a different letter).
 - Consider using a checklist where applicable to limit the need for entering text.
 - Include the ability to “uncheck” a rating if the evaluator inadvertently enters a score where none is wanted.
- User friendliness.
 - The scoring system feature for verifying data is necessary, but it would be more usable if it displayed only those lines where a rating was entered. This would allow the SME to quickly determine what evaluation had been entered.
 - Saving data should be an easy process.
- Improve operational use.
 - Make the scoring software password protected to prevent someone from accessing ratings and making changes without the evaluator’s knowledge.
 - Add links to doctrinal references for use as an evaluator refresher before doing the evaluation, similar to the “additional info” link in the initial scoring system.
 - The “alert” feature would be beneficial if the system were used with a pre-recorded session and could be used to notify the evaluator of an upcoming observation.
 - Create a means of identifying which scenario score sheet is being viewed. This is especially important when viewing/assessing several different scenarios within a short time, or when accessing previous ratings.

Revised Scoring System

Based on the try-out results, the team concluded that the core competencies were not as effective as needed for a scenario-independent system that has future application to other unit types and echelons. Therefore, the team reached a consensus that the competencies required major revision to improve both scenario independence and user friendliness in support of the project’s goals. The try-out observations indicated that the revised system must:

- Be usable while observing a scenario.
- Provide adequate detail to guide the evaluator.
- Avoid detail that detracts from evaluator performance.
- Reflect realistic revision priorities based on feasibility and impact.

- Be time sequenced in accordance with mission actions.
- Have an Army-relevant theoretical grounding.

The team developed a revised scoring system with ten competencies. The structure of the competencies stems from two sources: (a) the Army's established plan-prepare-execute (move, shoot, communicate)-consolidate/reorganize phases of mission accomplishment, and (b) Hiller's (2004) command and control (C2) information processing model incorporating SA and non-military factors.

Theoretical Basis – C2 Information Processing Model

Hiller's (2004) information processing model developed for C2 applications was used to frame the competencies (see Figure 3). This model led to the addition of two competencies (see Table 11, competencies #1 and #10) to the eight competencies which were based on the Army's mission accomplishment phases. Hiller emphasized diplomatic, informational, military, and economic (DIME) factors as critical for mission accomplishment. The DIME framework led the team to identify six specific factors at work at the squad level: religious interests, intelligence collection/generation, socio-cultural interests, civil affairs and infrastructure, attention grabbing potential, and military factors (RISCAM). Given the importance of RISCAM factors in both the contemporary operational environment (COE) and future training efforts, the team decided that definitive consideration of these factors in both the first and last tasks was essential.

Hiller's (2004) cognitive information-processing model (Figure 3) entails six overlapping and iterative phases:

1. Establish the Goal and Objectives. Acquire and understand the mission and analyze the situation to prioritize the desired effects (DIME) to support mission accomplishment. In short, set the Goal and objectives (**G**). This G function is then instrumental in guiding the selective review of existing data/information and collection and analysis of new data and information.
2. Conduct a preliminary review of the situation information and logically and intuitively form a "picture" from selective samples of data/information – intuitively, since there is potentially too much information and no hard and fast rules for distilling the information available into a rigorously derivable summary "picture." Thus, intuit the picture (**I**).
3. Continually review new information on a selective basis (e.g., intelligence summaries from various echelons, commercial news broadcasts) to discipline intuitive components of the picturing process and update the picture. Review and Adjust (**RA**) the picture based on new information.
4. Decide (**D**) on the course of action (COA), typically in collaboration with higher, parallel, and subordinate organizations, with review and approval by higher headquarters (HQ) as time and risk permit.
5. Command and control (**C2**). Issue mission orders and control, as appropriate.
6. Assess effects, which is actually intrinsic to C2, but separately identified in the model because of its importance as feedback for G and RA, as well as C2.

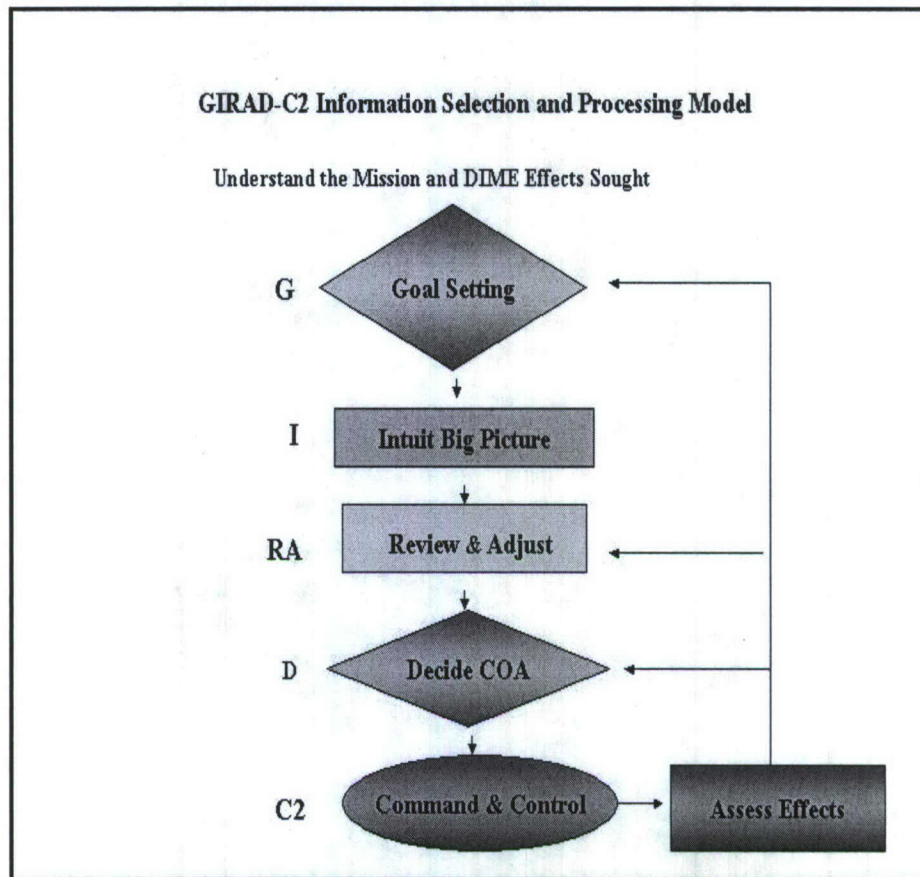


Figure 3. Hiller's C2 information processing model.

Characteristics of the Revised Scoring System

The revised scoring scheme contains ten competencies (performance criteria) at Level 1, with four to seven supporting actions (akin to knowledge, skills, and abilities) for each competency at Level 2 (Table 11). Also included in the scoring system are prompts for evaluator comments (see Figure 2). Each competency and each supporting action is scored on a five-point normative scale. The ratings for the supporting actions along with the comments can inform but not constrain the evaluator's rating of the parent competency. Overall scoring is achieved by computing the average number of points across the ten competencies, retaining the five-point scale as the interpretive context. To account for "not applicable" (NA) cases, the aggregation process computes the average using only the rated competencies. A weighting scheme could also be used, in which case weights and computational rules would need to be defined.

Table 11.

Revised Competencies with Supporting Actions

Competency	Supporting Action
1. Did the leader establish visualization of the battlefield and relate the mission to it?	<ul style="list-style-type: none"> • Acquire and review latest RISCAM considerations against Operation Order (OPORD) and commander's (Cdr's) intent? • Form "big picture" and include it as background for unit's OPORD? • Check validity of leader's intuition when forming COA, time permitting? • Explicitly consider adversary intent and likely reactions and counteractions in COA? • Consider RISCAM effects of mission execution in forming COA?
2. Did the leader conduct appropriate planning?	<ul style="list-style-type: none"> • Follow the 1/3 - 2/3 rule? • Issue subordinate leaders a Warning Order (WARNO)? • Provide guidance to subordinates in preparation for the OPORD? • Consider and properly use all available assets? • Issue timely, complete, and clear OPORD?
3. Did the leader/unit prepare appropriately?	<ul style="list-style-type: none"> • Conduct appropriate and sufficient rehearsal? • Adjust plan based on results of rehearsal and/or updated information? • Ensure all supporting assets are properly prepared? • Inspect and check equipment and Soldiers? • Begin mission at required time?
4. Did the leader/unit use the appropriate movement technique?	<ul style="list-style-type: none"> • Base their movement on probability of enemy contact? • Follow the planned route or identify appropriate reason to deviate? • Keep subordinates and higher informed of any changes? • Use proper movement control measures? • Maintain control of unit during entire movement? • Follow appropriate procedures at danger areas? • Meet all time requirements during movement?
5. Did the leader/unit react quickly and appropriately to enemy contact?	<ul style="list-style-type: none"> • Understand the type of contact? (sniper, improvised explosive device [IED], etc.) • Maintain positive control (maneuver and fires) during contact? • Provide Situation Report (SITREP) to higher and keep subordinates informed? • Handle wounded and prisoners of war (POWs) appropriately? • Enforce the Rules of Engagement (ROE) and Rules of Interaction (ROI)?
6. Did the leader/unit react appropriately to unexpected situations?	<ul style="list-style-type: none"> • Assess the situation? • Keep subordinates and higher informed? • Respond correctly based on circumstances? • Adjust remainder of mission based on revised situation?
7. Did the leader/unit execute proper actions on the objective?	<ul style="list-style-type: none"> • Begin actions at required time? • Act according to plan or as adjusted en-route? • Keep subordinates and higher informed of events? • Adjust actions based on evolving circumstances? • Comply with intent of higher? • Accomplish the mission?

Table 11.

Revised Competencies with Supporting Actions (continued)

<p>8. Did the leader/unit properly and adequately consolidate and reorganize?</p>	<ul style="list-style-type: none"> • Request an Ammunition, Casualty and Equipment (ACE) report from subordinate leaders? • Send higher a timely, complete and <u>accurate</u> SITREP? • Handle casualties and POWs appropriately? • Fill all key positions? • Recover and distribute key equipment? • Prepare for follow-on mission?
<p>9. Did the leader best employ all available assets?</p>	<ul style="list-style-type: none"> • Recognize all assets that were available? • Coordinate use of all assets? • Monitor actual use of all assets, and adjust as required? • Employ appropriate assets under proper circumstances?
<p>10. Did the leader update visualization of the battlefield and apply to new/revised mission?</p>	<ul style="list-style-type: none"> • Assess RISCAM effects of unit's actions? • Report to higher HQ, and receive mission updates or revised OPOD? • Acquire and review latest RISCAM considerations against OPOD and Cdr's intent? • Reform big picture and relate current mission to it for unit's next OPOD? • Adjust COA with attention to RISCAM effects?

User interface characteristics. The appearance and characteristics of the interface for the revised scoring system are generally the same as originally designed. However, the revised system incorporates the following characteristics:

- The tool's overview is expanded to explain the origin of the competencies, their applicability, and the reason for the five-point rating scale.
- All ten competencies (Level 1) and all supporting actions (Level 2) appear on the interface screen in a fixed order. Competencies and actions are numbered, with no shorthand tags. The five-point rating scale, plus "Not Applicable" (NA), accompanies each competency and each supporting action.
- Since the acronym RISCAM was developed for this project, wherever RISCAM appears a pop-up definition appears upon cursor hover to clarify its meaning to the user.
- A text box is accessible for each write-in prompt. Clicking a "Comments" icon causes a pop-up box to appear. Each box is tagged so it is clear which item the comments address.

Software considerations. Substantive suggestions for improving the scoring system software were collated and then evaluated. The deliberative process considered priorities of potential software enhancements in light of their importance and feasibility within the scope of the current project. The following changes were made to improve the ease of use:

- An explanation of the process to save ratings was included in both an electronic file and the scoring system overview.
- The blinking arrowhead was changed to a color that stands out better against the background.
- The scroll bar on the right side of the "record measures and comments" section was changed to a color that stands out better against the background.

- The reduction of the rated items in the revised scoring system (vs. the try-out version) allowed the revised scoring system to have one scrollable page for recording evaluations. This one page contains all core competencies and supporting behaviors and will greatly simplify use of the scoring system by an evaluator.

Although not feasible within the scope of this project, future software versions should include the following as high priorities:

- Text entry capability improvements to facilitate quick and easy entry of written comments while using a hand-held device. Use of voice recognition software may be a solution.
- Simplification of the process to save ratings.
- Development of those features which are notional in the prototype, including access to references and “alerts” functionality.

Lessons Learned

Key lessons learned which have applicability to subsequent work on training evaluation tools are discussed below.

Scenario/mission independence. Achieving scenario/mission independence is a major challenge. The Army has a long history of using ARTEPs for training evaluations with emphasis on achieving detailed standards for specific missions. Training evaluations are often scenario dependent. As evidenced by try-out of the initial scoring system, the use of specific criteria becomes too voluminous and complicated to be used for evaluation during a short, quick-paced scenario and focuses the evaluator on details specific to the scenario versus scenario independent skills in terms that are transferable to other missions. The revised scoring system is a departure from this norm and is intended to be both scenario independent and applicable to other unit types and echelons. Development of this innovative evaluation/measurement tool required the departure from the common Army practice and the merging of multiple frameworks. The determination of the system’s efficacy and acceptability to operational users awaits examination in future projects.

Alternative frameworks for the competencies/tasks. The initial scoring system competencies/tasks (used in the try-out) were based on the framework of the ARTEP and MTP. Based on the feedback from the try-out, the revised system’s framework is based primarily on the Army’s plan-prepare-execute (shoot, move, communicate)-consolidate/reorganize model. Shorthand tags for each task were considered to help users relate tasks to either the framework used, or to an alternative measurement framework. The team decided that the risks of doing so outweigh the potential benefits, especially in providing a larger set of tasks/steps that would be more cumbersome for evaluator users, as was the problem with the first version of the system.

Inclusion of COE realities. Specific COE conditions (e.g., IEDs, checkpoints) were not included because of the intent to provide measures that are mission and situation independent. Evaluators or commanders can include evaluation of COE specific actions within the framework of the competencies that are in the revised system.

The importance of SA and RISCAM. The importance of SA and RISCAM dimensions are evidenced in many media, ranging from the CSA's training guidance to the daily television news reports. Even if not always expressed as such, SA has always been a major requirement for unit leaders at all echelons. However, the factors included in RISCAM (or another version using similar factors) have not been a concern of the dismounted infantry squad leader until recently. The recognition that RISCAM factors must be considered at every echelon has come during the COE and is a function of numerous factors including the relative independence of small unit leaders, the nature of current operations, and the unprecedented level of media presence. In today's operational environment, and therefore the training evaluation environment, consideration of RISCAM factors are imperative to SA by leaders at all echelons. The RISCAM factors are relevant to all scenarios, types of units and echelons. Therefore, they are included in the competencies/tasks.

Task-level rating scale. A binary Go-No Go scale is commonly used for Army training evaluations, and this was noted in the feedback from the tryout. However, because of the research nature of this effort, a five-point rating scale for tasks and behaviors is built into the scoring system instead of the Go-No Go or T-P-U scale. An untraditional scale is likely to draw criticism and resistance from Army trainers. If deemed necessary or preferable for operational use of the scoring system, conversion to a Go-No Go scale may be appropriate.

Designation of leader vs. unit responsibilities. The competencies specify the leader and/or unit as responsible for a task, while supporting actions do not specify leader versus unit member responsibilities. Some tasks apply mainly to leaders while others apply to leaders and units. Extending the designation process to supporting actions would unnecessarily increase the volume of the scoring scheme and downplay the importance of evaluator judgment.

Aggregation flexibility. Providing a more flexible scheme for aggregating scores was considered. However, building flexibility into the tool would have required extensive analysis, definition, and programming beyond the scope of this project. The development team placed significant emphasis on competencies that are both scenario independent and appropriate for other unit types and echelons. Therefore, aggregation to emphasize a specific competency or set of competencies, which may be appropriate for specific research or training purposes, can be accomplished through weighting, omitting a competency or supporting actions, or adding a supporting behavior unique to the scenario or purpose of the evaluation.

Doctrinal reference links. No links to doctrinal references were included in the revised scoring system since such links were beyond the scope of the current project. They may be added in future projects. However, the SME feedback indicated that reference materials would be valuable in preparing for evaluation and for reference both during and after a scenario. Therefore, the issue and approaches for implementation merit investigation.

User-friendly interface. The literature review indicated that user friendliness of the measurement system is a major design consideration. This was borne out in the try-out. Even when used in a research environment, the SMEs were adamant that the system must be easy to use if it is to be accepted by users, and not merely by researchers. Developing a user friendly measurement system is a challenge. As evidenced in their recommendations for improving the

system, the try-out SMEs defined a need for a self-contained hand-held device that can be used as easily as a paper instrument while moving with the training unit in a field environment. As the SMEs stated, it must make the user more effective and efficient in their job.

Demonstrating an incomplete system. There are major advantages of demonstrating an incomplete system, including significant time and costs savings for development of components not essential to the demonstration. However, there are also significant disadvantages which should be considered for future efforts. The inability to evaluate all components of the system is clear. Additionally, without previous employment of the system, it may be difficult to forecast which components will not be needed, and therefore notional, during the actual demonstration. Failure to accurately anticipate SME use will not only result in frustrating the SMEs but failing to demonstrate what may be important linkages.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The revised scoring system that incorporates the try-out results is a valuable contribution to the technology for measuring learning and performance in collective training exercises. The most useful innovation is the generality of the measures, migrating from ARTEP based measures to a system that has applicability to diverse missions, various unit types, and multiple echelons. Most previous systems of measuring unit performance, including the scoring system initially developed for this project, have been based on a specific ARTEP with tasks, steps and standards unique to a specific scenario, mission, unit type and/or echelon. Based on the feedback from the try-out of the initial system early in this project, the team adopted an innovative approach in a major refinement of the scoring system. The revised system has both a theoretical foundation in Army training research (Hiller, 2004) and alignment with the Army's long established phases of mission accomplishment – plan-prepare-execute-consolidate/reorganize, rather than ARTEP specifics. The result is a system that follows the common scenario phases and applies to multiple scenarios, unit types and echelons.

Further research is needed to demonstrate the revised scoring system and ascertain the extent to which it is scenario independent and applicable to multiple unit types and echelons. Another major focus of the demonstration should be the extent to which users consider the evaluator tool usable in evaluating unit performance.

The research in this project yielded a number of lessons regarding the design and development of performance measurement tools. The following paragraphs summarize key lessons learned.

User friendliness is critical. If a scoring system is to be accepted for use operationally, as opposed to research only, the system must be perceived as easy to use, effective and efficient for measuring unit performance. This involves two critical aspects: a) design and structure of the scoring system and b) software design characteristics. Simply stated, if a scoring system is to be used, it must demonstrate that it enables the user to perform their duties more efficiently and effectively. Future research should consider these aspects and any demonstration should include specific measures of user friendliness along both dimensions.

Scoring system detail can be a detractor. A major issue with this project's initial ARTEP-based scoring system was that it contained too much detail. The detail occurred in the form of task steps and performance measures to support evaluation of a squad's performance. Although the detail was intended to help the evaluator, it detracted from the system's effectiveness and utility. Feedback from the try-out indicated that the large amount of detail narrowed the evaluator's focus to ARTEP specifics, resulting in three separate but interrelated problems: (a) it distracted the evaluators from their observation of the mission execution; (b) it focused attention on the details of the tasks being performed rather than broader aspects of performance; and (c) it led the evaluators to defer entering scores and comments into the system until mission execution ended. Therefore, limiting not only the number of core competencies but also the number of supporting actions is critical. Shorthand tags can help evaluators zero in on

key elements, but they cannot fully offset excessive detail. Finally, focusing core competencies and supporting actions on aspects which are common to multiple scenarios is critical for scenario independence.

A literature review can contribute. Applying lessons from the literature on measuring unit training performance can facilitate refinement of the scoring system during future research. Much of the literature stems from systems or environments that are specific to the scenario, unit type, or echelon. However, it appears that some of the lessons gleaned from the literature review in this project (Table) can foster a scoring system that is scenario independent and applicable to multiple unit types and echelons. Among the lessons are those pertaining to development of competencies, performance measurement, observer considerations, and observer job aids. The ones that proved critical to this project are listed in Table 12.

Table 12.
Critical Lessons from Literature Review

Category	Lesson Learned
Development of Competencies	<ul style="list-style-type: none"> – Detailed analysis is critical, including involvement of qualified SMEs – The analytical process should focus on key (vs. all) tasks or events – Research and real-world factors should be balanced – Development schedule should accommodate multi-phase, iterative process – Mission essential competencies offer a high-value approach
Performance Measurement	<ul style="list-style-type: none"> – Collective measures should reflect overall unit performance – Performance standards should reflect bands or ranges rather than point values – Measurement scaling can enhance differentiation of performance – Feasibility (via automated or evaluator collection) is a key criterion – Design of measures should consider output format requirements
Observer Considerations	<ul style="list-style-type: none"> – Capabilities of target audience evaluators should be defined – Highly proficient, motivated evaluators should be a key priority – Managing evaluator workload is a central consideration
Observer Job Aids	<ul style="list-style-type: none"> – Evaluators must maintain good SA – Ready access to references can enhance evaluator effectiveness

System design must leverage multi-disciplinary expertise. Using a design team that combines the knowledge and skills of both SMEs and behavioral scientists in a collaborative exchange of views can benefit the process and outcome of the project. One of the literature review lessons learned for developing the competencies was that a detailed analysis is critical and must include SMEs. This was reinforced in this project. But a major lesson from the two design efforts of this project was that the SMEs and behavioral scientists must participate interactively to ensure all factors are considered and integrated.

Qualifications of the team's SMEs are critical. The literature review correctly indicated that integrating well qualified SMEs into the research team is critical. The depth of expertise and professional judgment of this project's SMEs played a critical role in ensuring the credibility of the scoring scheme and the quality of the try-out feedback. This aspect of development and demonstration should receive considerable emphasis.

Recommendations

The development of a scenario independent performance measurement system would provide Army leaders and trainers a valuable tool to determine if a training intervention did in fact produce learning. Measurement of learning during training of units using multiple scenarios would no longer be hampered by the circumstance that units seldom repeat scenarios. Leaders and trainers would be able to conclusively measure the performance of Soldiers as they apply lessons learned during one scenario while executing subsequent scenarios.

This report provides the conceptual framework for measuring and interpreting unit performance across diverse scenarios and training conditions. The following recommendations can facilitate application of the performance assessment approach to develop its potential as a training tool:

- Try out the revised scoring system to determine scenario independence and applicability to other unit types and echelons and to provide data for further improvement.
- Inform Army leaders of this innovative approach to measuring unit collective performance and solicit their support in evaluating the scoring system.
- Gather user feedback for use by follow-on investigators who continue to refine the scoring system.
- Further investigate methods and techniques that will produce a user interface that maximizes evaluator efficiency and effectiveness.

This research has application to a wide range of unit types and echelons and contributes to the body of knowledge on Army training. This is a critical contribution to improving the combat effectiveness of Army units, especially in the COE which necessitates training to deal with disparate mission sets across the spectrum of conflict in any environment, under all conditions.

REFERENCES

- Alliger, G., Garrity, M.J., Morley, R.M., McCall, J.M., Beer, L., & Rodriguez, D. (2003). *Mission essential competencies for the AOC: A basis for training needs analysis and performance improvement*. Paper presented at the annual Interservice/Industry Training, Simulation, and Education Conference.
- Colegrove, C.M., & Bennett, W., Jr. (2006). *Competency-based training: Adapting to warfighter needs*. Mesa, AZ: Air Force Research Laboratory, Human Effectiveness Directorate.
- Fowlkes, J.E., Lane, N.E., Salas, E., Franz, T. & Oser, R. (1994). Improving the measurement of team performance: The TARGETS methodology. *Military Psychology*, 6(1), 47-61.
- Casey, G. W., Jr. (2007). The strength of the nation. *Army*, 57(10). 19-30.
- Hiller, J.H. (2004). *Proposed comprehensive joint operations performance measurement architecture for achieving the Joint Assessment and Enabling Capability (JAEC) at level 1* (Unpublished Report). Albuquerque, NM: Northrop Grumman Corporation.
- Johnston, J.H., Vincenzi, D.A., Radtke, P.H., Salter, W., & Freeman, J. (2005). *A human systems integration method for validating team performance assessment within a simulation-based training system*. Paper presented at the Human Factors and Engineering Society meeting.
- Knerr, B.W., Lampton, D.R., Thomas, M., Comer, B.D., Grosse, J.R., Centric, J.H., Blankenbeckler, P., Dlubac, Wampler, R.L., M., Siddon, D., Garfield, K., Martin, G.A., & Washburn, D.A. (2003). *Virtual environments for dismounted soldier simulation, training, and mission rehearsal: Results of the FY 2002 culminating event* (ARI Technical Report 1138). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (ADA417360)
- Kyne, M.M., Militello, L.G., Thordsen, M.L. & Klein, G. (2002). *Teamwork assessment scales for C2 functions of battalions and brigades* (ARI Research Note 2002-08). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (ADA400488)
- Kyne, M.M., Thordsen, M.L., & Kaempf, G. (2002). *A model-based team decision-making and performance assessment instrument: Development and evaluation Volumes I and II* (ARI Research Note 2002-09). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (ADA400491)
- Throne, M.H., Holden, W.T., Jr., & Lickteig, C.W. (2000). *Refinement of prototype staff evaluation methods for future forces: A focus on automated measures* (ARI Research Report 1764). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (ADA384027)

APPENDIX A

Acronyms and Abbreviations

AAR	After Action Review
ACE	Ammunition, Casualty and Equipment
ARI	U.S. Army Research Institute for the Behavioral and Social Sciences
ARTEP	Army Training and Evaluation Program
C2	Command and Control
CBRN	Chemical, Biological, Radiological, and Nuclear
CD	Compact Disc
COA	Course of Action
COE	Contemporary Operational Environment
CSA	Chief of Staff of the Army
D	Decide COA
DIME	Diplomatic, Informational, Military and Economic
DIVAARS	Dismounted Infantry Virtual After Action Review System
E7	Sergeant First Class
E9	Sergeant Major
EPW	Enemy Prisoner of War
FRAGO	Fragmentary Order
G	Goal Setting
HQ	Headquarters
html	Hyper Text Markup Language
I	Intuit Big Picture
IAW	in accordance with
IED	Improvised Explosive Device
JAEC	Joint Assessment and Enabling Capability
LD	Line of Departure
Ldr	Leader
MEC	Mission Essential Competencies
METL	Mission Essential Task List
METT-TC	Mission, Enemy, Terrain, Troops, Time Available, and Civil Considerations
MLPCTE	Measuring Learning and Performance in Collective Training Exercises
MOUT	Military Operations on Urban Terrain
MTP	Mission Training Plan
NA	Not Applicable
NCO	Non-Commissioned Officer
O4	Major
O/C	Observer Controller
OPFOR	Opposing Force
OPORD	Operation Order
OPSEC	Operation Security
POSNAV	Positioning and Navigation system
POW	Prisoner of War

RA	Review and Adjust
REDCON	Readiness Condition
RISCAM	Religious interests, Intelligence collection/generation, Socio-cultural interests, Civil affairs and infrastructure, Attention grabbing potential, and Military factors
ROE	Rules of Engagement
ROI	Rules of Interaction
SA	Situational Awareness
SITREP	Situation Report
SME	Subject Matter Expert
SU	Situational Understanding
TACSOP	Tactical Standing Operating Procedures
TARGETS	Target Acceptable Responses to Generated Events or Tasks
TBTRU	Technology-Based Training Research Unit
TLP	Troop Leading Procedures
T-P-U	Trained, Needs Practice, Untrained
TSP	Training Support Package
TTP	Tactics, Techniques and Procedures
UN	United Nations
WARNO	Warning Order

APPENDIX B

Data Collection Instruments

SME WORKSHEET

INSTRUCTIONS: SMEs will use this worksheet to record their reactions as they operate the O/C workstation. A worksheet is to be completed at the end of every exercise.

1. What is the name of the scenario? _____
2. When did the exercise run? Start Time _____ Stop Time _____
3. Which task(s) did you use to record performance in this exercise?
4. How well did the steps and sub-steps fit the actions and performance of the squad? Were there too many or too few steps and sub-steps?
5. Describe any problems you encountered in measuring all aspects of squad performance.
6. How well could you keep up with the pace of the exercise? Describe any problems.
7. Did you have any problems with navigating, reading (legibility), scrolling, entering comments, etc.? Please describe.
8. How would you improve the scoring system? Consider tasks, organization, level of detail, 5-point scale, etc.
9. How would you improve the user interface? Consider menus, legibility, ease of use, knowing where you are, presentation of text, distinctiveness of elements, text entry, etc.
10. Other comments?

OBSERVATION GUIDE

INSTRUCTIONS: The try-out O/C will use this form to record administrative information and observations. The form covers all train-up and scoring activities. It may be necessary to use supplemental sheets to record detailed observations.

Familiarization and Train-up

1. Enter the information for each practice exercise in the table below.

Name of Practice Scenario	Start Time	Stop Time
1.		
2.		
3.		

2. What were the big challenges for the SMEs as they learned to operate the O/C workstation?
3. What questions and issues did the SMEs raise during train-up?
4. What substantive comments did the SMEs make as they practiced?
5. Did you make any adjustments to the train-up plan as things unfolded? If yes, describe.
6. How proficient were the SMEs when they finished train-up? Were you satisfied?
7. Did any technical problems arise during train-up? If yes, describe.

Scoring Session I

8. Enter the information for each exercise scored by the SMEs.

Name of Scenario	Start Time	Stop Time	Replays?
1.			
2.			
3.			
4.			
5.			

9. Did the scenarios provide a reasonable test of the scoring system? If not, explain.
10. Did you see any signs of concern or confusion as the SMEs scored the exercises? Explain.
11. Did the SMEs compare notes between exercises? If yes, describe.
12. How did the SMEs differ in their approach to assessing squad performance?

13. When did the SMEs choose to replay part of a scenario? Why?
14. How often did the SMEs continue their scoring after an exercise ended? For how long?
15. What questions and issues did the SMEs raise during or between exercises?
16. What substantive issues did you resolve, and how?
17. Did you make any adjustments to the scenario execution plan? If yes, describe.
18. What substantive comments did the SMEs make during this session?
19. Did any technical problems arise during the session? If yes, describe them.

Scoring Session II

20. Enter the information for each exercise scored by the SMEs.

Name of Scenario	Start Time	Stop Time	Replays?
6.			
7.			
8.			
9.			
10.			

21. Did the scenarios provide a reasonable test of the scoring system? If not, explain.
22. Did you see any signs of concern or confusion as the SMEs scored the exercises? Explain.
23. Did the SMEs compare notes between exercises? If yes, describe.
24. How did the SMEs differ in their approach to assessing squad performance?
25. When did the SMEs choose to replay part of a scenario? Why?
26. How often did the SMEs continue their scoring after an exercise ended? For how long?
27. What questions and issues did the SMEs raise during or between exercises?
28. What substantive issues did you resolve, and how?
29. Did you make any adjustments to the scenario execution plan? If yes, describe.
30. What substantive comments did the SMEs make during this session?
31. Did any technical problems arise during the session? If yes, describe them.

HOTWASH GUIDE

INSTRUCTIONS: The try-out coordinator will use the questions in this form to guide discussion during the hotwashes. Other questions will no doubt arise. The focus will likely shift from one hotwash to the next. Be sure to work in the special concerns and hot buttons of the SMEs.

1. Did you get enough training on the workstation and scoring procedures? How proficient did you feel at the end of the train-up?
2. How well does the scoring system fit the operations of dismounted infantry squads?
3. To what extent are the tasks and steps doctrinally correct? What needs to be changed?
4. How well would the scoring system fit platoon operations? Company operations?
5. How easy was it to find a particular task or step quickly?
6. Did you feel like you always knew where you were in the maze of tasks and steps? Did you ever feel lost or disoriented?
7. What significant problems did you encounter with the scoring system? How did you adjust your scoring activities because of the problems?
8. Did the workstation make it easier to measure squad performance? Explain.
9. How often did you enter remarks or comments about performance? How easy was it to enter comments?
10. Did you have any problems keeping up with the pace of the exercise?
11. How often did you want to revisit ratings or comments you recorded earlier? How easy did the workstation make it for you to do so? How easy was it to change things?
12. Does the workstation give you every capability you need to *measure* performance? What falls short? What's missing?
13. Did the O/C workstation cause you to do things differently than you normally would? How so?
14. Is the workstation interface user friendly and easy to operate? How would you improve the interface?
15. Did the menu structure make it easy for you to navigate through the workstation functions? What problems did you encounter? How would you improve the menu features?
16. Did the workstation provide enough cues to tell you where you were and how you got there? How would you improve these cues?

17. What significant problems did you encounter with the user interface? How did they impact your scoring activities?
18. Did you use the workstation differently in the last exercise compared to the first exercise? Explain.
19. How could you use the O/C tool to track a unit's proficiency over several exercises? What common measures would you focus on?
20. Do infantry trainers need a portable O/C tool such as this? Why?
21. Do you think the assessment tool would help O/Cs train infantry small units? How so?
22. Would the assessment tool make the job of measuring performance easier? Elaborate.
23. Could the assessment tool help standardize performance measurement? Explain how and why.
24. Would a portable tool help O/Cs prepare for and conduct after action reviews (AARs)? Explain.
25. Do you think infantry trainers would use a portable O/C tool? Would they have to change their mindset a lot? Elaborate.
26. What training would an O/C need to use the assessment tool effectively?
27. What obstacles might stand in the way of infantry trainers using a portable O/C tool? Can these concerns be resolved successfully?
28. The assessment tool would need more capabilities to support exercise preparation, performance feedback, AARs, archiving, etc. What additional capabilities are essential?
29. To what extent did the scenarios reflect current doctrine and Tactics, Techniques and Procedures (TTP)? What needs to be changed?
30. How well did the scenarios portray infantry squad operations? How would you improve them?
31. Did the scenario-driven events and actions provide a good test of the O/C workstation's capabilities? Explain.
32. Did the Dismounted Infantry Virtual After Action Review System (DIVAARS) workstation provide an effective environment for simulating squad operations? Elaborate.
33. How adequate were the conditions for testing the O/C workstation capabilities? Consider the scope of operations, tactical realism, number of exercises, time available, flexibility of observation, workarounds, technical problems, etc.
34. Any other thoughts or comments?

APPENDIX C

Literature Review References

- Alliger, G., Garrity, M.J., Morley, R.M., McCall, J.M., Beer, L., & Rodriguez, D. (2003). *Mission essential competencies for the AOC: A basis for training needs analysis and performance improvement*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference.
- Archer, R., Walters, B., Oster, A., & Van Voast, A. (2002). *Improving soldier factors in prediction models* (ARI Technical Report 1132). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Bell, H.H., Dwyer, D.J., Love, J.F., Meliza, L.L., Mirabella, A., & Moses, F.L. (1997). *Recommendations for planning and conducting multi-service tactical training with distributed interactive simulation technology* (ARI Research Product 97-03). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (ADA38480)
- Bennett, W., Schreiber, B., & Andrews, D. (2002). Developing competency-based methods for near-real-time air combat problem solving assessment. *Computers in Human Behavior*, 11(6), 773-782.
- Boldovici, J., Bessemer, D., & Bolton, A. (2002). *The elements of training evaluation* (ARI Book). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Colegrove, C.M., & Bennett, W., Jr. (2006). *Competency-based training: Adapting to warfighter needs*. Mesa, AZ: Air Force Research Laboratory, Human Effectiveness Directorate.
- Denning, T., Bennett, W., Bell, J., & Landrum, L. (2003). *Tactics development and training program validation in distributed mission training: A case study and evaluation with the USAF weapons school*. Paper presented at the Interservice/Industry Training, Simulation and Education Conference.
- Fowlkes, J.E., Lane, N.E., Salas, E., Franz, T. & Oser, R. (1994). Improving the measurement of team performance: The TARGETS methodology. *Military Psychology*, 6(1), 47-61.
- Furman, J.S., & Wampler, R.L. (1982). *Methodology for evaluating unit tactical proficiency at the National Training Center* (Master's Thesis). Monterey, CA: Naval Postgraduate School. (ADA119574)
- Gately, M., Watts, S.M., & Pleban, R.J. (2002). *Assessing decision-making skills in virtual environments* (ARI Technical Report 1130). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (ADA405079)

- Gehr, S., Schreiber, B. & Bennett Jr., W. (2004). *Within-simulator training effectiveness evaluation*. Paper presented at the Interservice/Industry Training, Simulation and Education Conference.
- Gehr, S., Schurig, M., Jacobs, L., van der Pal, J., Bennett, W., Jr., & Schreiber, B. (Undated). *Assessing the training potential of MTDS in exercise first wave RTO-MSG-35, pp. 12-1 – 12-15. NATO R&T Organization*.
- Holden, W. T., Jr., Throne, M.H., & Sterling, B.S. (2001). *Prototype automated measures of command and staff performance* (ARI Research Report 1779). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (ADA397634)
- Johnston, J.H., Vincenzi, D.A., Radtke, P.H., Salter, W., & Freeman, J. (2005). *A human systems integration method for validating team performance assessment within a simulation-based training system*. Paper presented at the annual conference of the Human Factors and Engineering Society.
- Kyne, M.M., Militello, L.G., Thordsen, M.L., & Klein, G. (2002). *Teamwork assessment scales for C2 functions of battalions and brigades* (ARI Research Note 2002-08). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences. (ADA400488)
- Kyne, M.M., Thordsen, M.L., & Kaempf, G. (2002). *A model-based team decision-making and performance assessment instrument: development and evaluation Volumes I and II* (ARI Research Note 2002-09). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (ADA400491)
- Lampton, D.R., Riley, J.M., Kaber, D.B., Sheik-Nainar, M.A., & Endsley, M.R. (Undated). *Use of immersive virtual environments for measuring and training situation awareness*. Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.
- Leibrecht, B.C. (1997). *An integrated database of unit training performance: Description and lessons learned* (ARI Contractor Report 97-25). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (ADA328669)
- Leibrecht, B.C., Kerins, J.W., Ainslie, F.M., Sawyer, A.R., Childs, J.M., & Doherty, W.J. (1992). *Combat vehicle command and control systems: Simulation-based company-level evaluation* (ARI Technical Report 950). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (ADA251917)
- Leibrecht, B.C., Lockaby, K.J., Perrault, A., Strauss, C., & Meliza, L.L. (2006). *Delivery of digital measurement guidance (D2MG)*. Killeen, TX: Northrop Grumman Mission Systems.

- Lewman, T., Mullen III, T., & Root, J. (1994). A conceptual framework for measuring unit performance. In Holz, R.F., Hiller, J.H., & McFann, H.H. (Eds.), *Determinants of effective unit performance: Research on measuring unit training readiness* (pp. 17-38). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Lickteig, C., Sanders, W., Durlach, P., & Lussier, J. (2003). *Human performance essential to battle command: Report on four future combat systems command and control experiments* (ARI Research Report 1812). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Meliza, L.L., Bessemer, D.W., Burnside, B.L., & Shlechter, T.M. (1992). *Platoon-level after action review aids in the SIMNET unit performance assessment system* (ARI Technical Report 956). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (ADA254904)
- Pleban, R.J., & Salvetti, J. (2003). *Using virtual environments for conducting small unit dismounted mission rehearsals* (ARI Research Report 1806). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Thordsen, M., Kyne, M., & Klein, G. (2002). *A model of advanced team decision making and performance: Summary report* (ARI Research Note 2002-10). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Throne, M.H., Holden, W.T., Jr., & Lickteig, C.W. (2000). *Refinement of prototype staff evaluation methods for future forces: A focus on automated measures* (ARI Research Report 1764). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (ADA384027)

APPENDIX D

Details of Core Competency Tasks As Used During Try-out and Subsequently Rejected

Summary of Tasks and Steps within Each Task

Task	Steps
Breach an Obstacle	<ul style="list-style-type: none"> • Leader (Ldr) gains and or maintains situational understanding (SU) • Unit determines nature of enemy obstacle • Ldr or designated representative reports obstacle to higher headquarters • Ldr decides to bypass or breach the obstacle • Ldr plans using troop-leading procedures • Ldr request additional resources based on the Mission, Enemy, Terrain, Troops, Time Available, and Civilian Considerations (METT-TC) • Ldr disseminates information to unit members to keep them abreast of the situation • Ldr issues orders and instructions • Unit conducts a rehearsal • Ldr issues a fragmentary order (FRAGO), as necessary • Unit executes the bypass • Unit executes the breach • Ldr reports completion of the bypass or breach to higher headquarters • Unit consolidates and reorganizes as necessary • Unit secures enemy prisoners of war (EPWs) if applicable • Unit treats and evacuates casualties as required • Unit processes captured documents and or equipment if applicable • Unit continues operations as ordered
Conduct a Defense	<ul style="list-style-type: none"> • Ldr gains and or maintains SU • Ldr receives an operations order (OPORD) or FRAGO and issues warning order (WARNO) • Ldr plans using troop leading procedures • Ldr or designated representative coordinates with adjacent unit as required • Ldr disseminates information to unit members to keep them abreast of the situation • Ldr briefs scheme of maneuver • Ldr issues orders and instructions • Unit conducts a rehearsal • Ldr issues FRAGOs, with changes to plan based on the rehearsal • Unit starts movement to a tactical assembly area (AA) or designated area short of the defensive position(s) • Ldr and recon element conducts the recon based on METT-TC • Ldr adjust the plan based on updated intelligence and recon effort • Ldr updates the enemy situation and disseminates • Unit moves tactically to assigned defensive positions

Summary of Tasks and Steps within Each Task

Task	Steps
	<ul style="list-style-type: none"> • Unit improves fighting positions • Ldr consolidates sketch cards and finalizes the unit fire plan • As time permits, Ldr directs improvement of positions • Ldr adjusts readiness condition (REDCON) level and disseminates • Unit conducts the defense • Unit consolidates and reorganizes as necessary • Unit secures EPW as required • Unit treats and evacuates casualties • Unit processes captured documents and or equipment as required • Unit continues operations as directed
Conduct a Movement to Contact	<ul style="list-style-type: none"> • Ldr gains and or maintains SU • Ldr receives an OPORD or FRAGO and issues WARNO • Ldr plans using troop leading procedures • Ldr disseminates information to each squad • Ldr prepares for the movement to contact • Ldr issues orders and instructions to include rules of engagement (ROE) and rules of interaction (ROI) • Unit conducts a rehearsal • Ldr issues FRAGOs, as necessary, for changes based on the rehearsal • Unit enters waypoints into positioning and navigation system (POSNAV) equipment to aid navigation • Unit uses approach march technique, based on the enemy situation • Unit consolidates and reorganizes as necessary • Unit secures EPWs as required • Unit treats and evacuates casualties • Unit processes captured documents and or equipment as required • Unit continues operations as directed
Conduct a Security Patrol	<ul style="list-style-type: none"> • Ldr gains and or maintains SU • Ldr receives an OPORD or FRAGO and issues WARNO • Ldr plans using troop leading procedures • Ldr issues orders and instructions to include ROE and ROI • Unit conducts rehearsal • Ldr issues FRAGOs, as necessary, to address changes to the plan based on rehearsal • Unit conducts security patrol • Unit consolidates and reorganizes as necessary • Unit secures EPWs as required • Unit treats and evacuates casualties • Ldr completes the patrol report • Unit processes captured documents and or equipment as required

Summary of Tasks and Steps within Each Task

Task	Steps
Conduct Tactical Movement	<ul style="list-style-type: none"> • Unit continues operations as directed • Ldr gains and or maintains SU • Ldr receives an OPORD or FRAGO and issues WARNO • Ldr plans using troop leading procedures • Ldr disseminates information to unit members to keep them abreast of the situation • Ldr briefs the movement plan • Ldr issues orders and instructions to include ROE and ROI • Unit conducts a rehearsal • Ldr issues FRAGOs, as necessary, to address changes to the plan based on rehearsal • Ldr and reconnaissance element conducts the reconnaissance METT-TC • Ldr adjusts the plan based on updated intelligence and reconnaissance effort • Ldr disseminates updated reports (if applicable), overlays, and other pertinent information • Ldr initiates movement to line of departure (LD) • Unit conducts passage of lines, if required • Unit moves using appropriate formation designated by Ldr • Unit executes movement technique as directed by Ldr • Ldr positions himself where he can best control and execute the desired formation • Unit maintains formation in accordance with (IAW) Ldr's guidance or TACSOP • Unit orients weapons and or weapon systems to provide security and maximize firepower • Unit moves undetected to the designated point specified in the OPORD • Unit consolidates and reorganizes as necessary • Unit secures EPWs as required • Unit treats and evacuates casualties • Unit processes captured documents and or equipment as required • Unit continues operations as directed
Conduct an Attack	<ul style="list-style-type: none"> • Ldr gains and/or maintains SU • Ldr receives an OPORD or FRAGO and issues WARNO • Ldr plans using troop leading procedures • Unit begins necessary movement • Ldr conducts a leader's reconnaissance • Ldr adjust the plan based on updated intelligence and reconnaissance effort • Ldr disseminates updated reports (if applicable), and other pertinent information • Unit prepares for attack • Unit issues FRAGOs as necessary , to address changes to the plan based on rehearsal • Unit executes the attack • Unit secures EPWs as required

Summary of Tasks and Steps within Each Task

Task	Steps
	<ul style="list-style-type: none"> • Unit conducts consolidation and reorganization • Ldr assesses and reports the situation to higher headquarters • Unit treats and evacuates casualties • Unit processes captured documents and or equipment as required • Unit continues operations as directed
Maintain Operations Security	<ul style="list-style-type: none"> • Ldr uses all information sources available to understand the tactical situation • Ldr protects friendly information; safeguards weapons, ammo and sensitive items • Unit employs active and passive security measures • Unit enforces litter discipline by collecting, securing, and disposing of trash securely • Ldr enforces radio discipline • Ldr enforces noise discipline
Action on Contact	<ul style="list-style-type: none"> • Ldr gains and or maintains SU • Unit deploys and reports • Unit complies with ROE and ROI • Ldr evaluates the situation • Ldr disseminates reports (if applicable), and other pertinent information to unit • Ldr selects an appropriate course of action (COA) based on all factors • Ldr uses cross talk with other units as necessary to obtain support • Ldr directs unit to execute COA based on the situation or commander's order • Ldr or subordinate keeps the commander informed throughout the operation • Unit consolidates and reorganizes as necessary • Unit handles EPWs if applicable • Unit treats and evacuates casualties if applicable • Unit processes captured documents and or equipment if applicable • Unit continues operations as directed
Conduct Troop-leading Procedures	<ul style="list-style-type: none"> • Ldr uses all information sources available to understand the tactical situation • Ldr receives an OPORD or FRAGO and issues WARNO • Ldr conducts mission analysis • Ldr makes a tentative plan • Ldr initiates movement IAW orders and or unit TACSOP • Ldr conducts reconnaissance • Ldr completes the plan • Ldr issues orders and instructions to include ROE and ROI • Ldr supervises preparations and refines the order

APPENDIX E

Details of Revised Infantry Squad Evaluation Criteria and Scoring

OVERVIEW: The revised scoring scheme contains 10 performance criteria (competencies, Level 1) and 4-7 supporting actions (akin to SKAs, Level 2) for each criterion, plus prompts for comments. Each competency and each supporting action will be scored on a five-point scale (normative). The ratings for the supporting actions along with the comments will inform but not constrain the evaluator's rating of the parent competency.

DEFINITION: RISCAM = religious interests, intelligence collection/generation, socio-cultural interests, civil affairs and infrastructure, attention grabbing potential, and military factors.

Competencies with Supporting Actions

1. Did the leader establish visualization of the battlefield and relate the mission to it?
 - a. Acquire and review latest RISCAM considerations against OPORD and Cdr's intent?
 - b. Form "big picture" and include it as background for unit's OPORD?
 - c. Check validity of leader's intuition when forming COA, time permitting?
 - d. Explicitly consider adversary intent and likely reactions and counteractions in COA?
 - e. Consider RISCAM effects of mission execution in forming COA?
 - f. Were significant mistakes made? Explain. [Write-in, unscored]
 - g. What were the best aspects of performance? [Write-in, unscored]
 - h. How might the training be improved? [Write-in, unscored]

2. Did the leader conduct appropriate planning?
 - a. Follow the 1/3 - 2/3 rule?
 - b. Issue subordinate leaders a WARNO?
 - c. Provide guidance to subordinates in preparation for the OPORD?
 - d. Consider and properly use all available assets?
 - e. Issue timely, complete, and clear OPORD?
 - f. Observer comments? [Write-in, unscored]

3. Did the leader/unit prepare appropriately?
 - a. Conduct appropriate and sufficient rehearsal?
 - b. Adjust plan based on results of rehearsal and/or updated information?
 - c. Ensure all supporting assets are properly prepared?
 - d. Inspect and check equipment and Soldiers?
 - e. Begin mission at required time?
 - f. Observer comments? [Write-in, unscored]

4. Did the leader/unit use the appropriate movement technique?
 - a. Base their movement on probability of enemy contact?
 - b. Follow the planned route or identify appropriate reason to deviate?
 - c. Keep subordinates and higher informed of any changes?
 - d. Use proper movement control measures?
 - e. Maintain control of unit during entire movement?
 - f. Follow appropriate procedures at danger areas?
 - g. Meet all time requirements during movement?
 - h. Observer comments? [Write-in, unscored]

5. Did the leader/unit react quickly and appropriately to enemy contact?
 - a. Understand the type of contact? (Sniper, IED, CBRN, ambush, etc.)
 - b. Maintain positive control (maneuver and fires) during contact?
 - c. Provide SITREP to higher and keep subordinates informed?
 - d. Handle wounded and POWs appropriately?
 - e. Enforce the ROE and ROI?
 - f. Observer comments? [Write-in, unscored]

6. Did the leader/unit react appropriately to unexpected situations?
 - a. Assess the situation?
 - b. Keep subordinates and higher informed?
 - c. Respond correctly based on circumstances?
 - d. Adjust remainder of mission based on revised situation?
 - e. Observer comments? [Write-in, unscored]

7. Did the leader/unit execute proper actions on the objective?
 - a. Begin actions at required time?
 - b. Act according to plan or as adjusted en-route?
 - c. Keep subordinates and higher informed of events?
 - d. Adjust actions based on evolving circumstances?
 - e. Comply with intent of higher?
 - f. Accomplish the mission?
 - g. Observer comments? [Write-in, unscored]

8. Did the leader/unit properly and adequately consolidate and reorganize?
 - a. Request an ACE report from subordinate leaders?
 - b. Send higher a timely, complete and accurate SITREP?
 - c. Handle casualties and POWs appropriately?
 - d. Fill all key positions?
 - e. Recover and distribute key equipment?
 - f. Prepare for follow-on mission?
 - g. Observer comments? [Write-in, unscored]

9. Did the leader best employ all available assets?
 - a. Recognize all assets that were available?
 - b. Coordinate use of all assets?

- c. Monitor actual use of all assets, and adjust as required?
 - d. Employ appropriate assets under proper circumstances?
 - e. Observer comments? [Write-in, unscored]
10. Did the leader update visualization of the battlefield and apply to new/revised mission?
- a. Assess RISCAM effects of unit's actions?
 - b. Report to higher HQ, and receive mission updates or revised OPORD?
 - c. Acquire and review latest RISCAM considerations against OPORD and Cdr's intent?
 - d. Reform big picture and relate current mission to it for unit's next OPORD?
 - e. Adjust COA with attention to RISCAM effects?
 - f. Were significant mistakes made? Explain. [Write-in, unscored]
 - g. What were the best aspects of performance? [Write-in, unscored]
 - h. How might the training be improved? [Write-in, unscored]

Rating Scale for Competencies (Level 1) and Supporting Actions (Level 2)

5	4	3	2	1	NA
Exceptional	Good	Average	Fair	Poor	Not Applicable

Overall Scoring

To compute a total overall score, compute the average number of points across the 10 competencies, retaining the five-point scale as the interpretive context. To account for NA cases, compute the average using only the rated competencies. (A weighting scheme could be used, in which case weights and computational rules would need to be defined.)

Example of unweighted overall score computation: If 8 of the 10 competencies were rated and the rating values were 4-2-3-3-2-4-3-4, then:

The total rated points would be 25.

The number of rated competencies would be 8.

The overall score (average) would be $25/8 = 3.12$.